Regression Analysis of United States Airfare

Sarah Hartness

Advisor: Ying-Ju Chen

Objective of the Research

Analyze airfare for the most frequent roundtrip, domestic flights in the US.

Goal:

Find what variables are important in determining airfare and pick four frequent, roundtrip flights to analyze

Why: Determining what variables affect flight price could encourage airlines to find ways to decrease the costs of those variables to make flying more accessible. In addition, determine if there is a seasonal effect on flight prices

How: Using a statistical model, we can find a regression to look at the significance of different independent variables in relation to the dependent variable, flight price

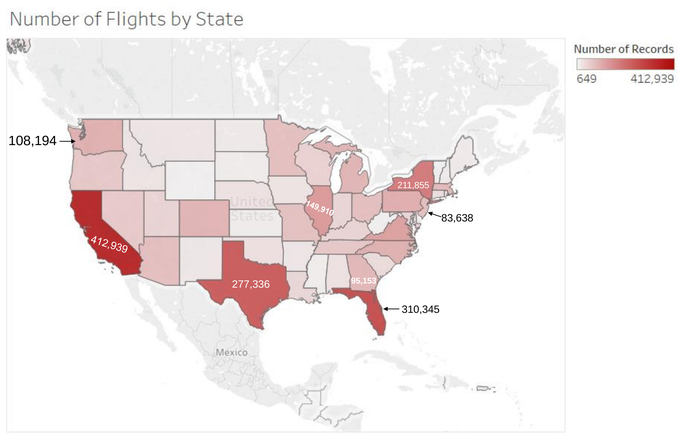
Summary of Data

- United States Department of Transportation - Bureau of Transportation Statistics
  - Airline Origin and Destination Survey 2017
  - Survey takes a 10% sample of airline tickets from reporting carriers.
  - The total number of records in the sample is 3,710,127
  - This project uses quarterly data (Quarter 1 – Quarter 4)
- All flights are roundtrip

Airport Codes and Cities

ATL	Atlanta, Georgia	LAX	Las Angeles, California
BOS	Boston, Massachusetts	LGA	New York City, New York
DEN	Denver, Colorado	MCO	Orlando, Florida
DFW	Dallas-Fort Worth, Texas	MSP	Minneapolis-St. Paul, Minnesota
DTW	Detroit, Michigan	ORD	Chicago, Illinois
EWR	Newark, New Jersey	PHX	Phoenix, Arizona
FLL	Fort Lauderdale, Florida	SEA	Seattle-Tacoma, Washington
JFK	New York City, New York	SFO	San Francisco, California
LAS	Las Vegas, Nevada		

Number of flights in 2017 by State



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Number of Records. Details are shown for Origin Country and Origin State Nm.

Most Frequent Roundtrip Flights

Quarter 1			Quarter 2			Quarter 3			Quarter 4		
Origin	Destination	Frequency	Origin	Destination	Frequency	Origin	Destination	Frequency	Origin	Destination	Frequency
ORD	LGA	3,213	LAX	JFK	3,615	LAX	JFK	3,406	LAX	JFK	3,970
JFK	LAX	3,174	JFK	LAX	3,206	JFK	SFO	3,391	JFK	LAX	3,678
LAX	JFK	2,786	ORD	LGA	3,200	EWR	SFO	3,160	ORD	LGA	3,224
SEA	LAX	2,560	EWR	SFO	2,953	ORD	LGA	2,805	SEA	LAX	2,922
EWR	SFO	2,510	SFO	BOS	2,735	SFO	SEA	2,664	EWR	SFO	2,872
BOS	MCO	2,453	ATL	LGA	2,626	BOS	SFO	2,612	ATL	LGA	2,782
MSP	PHX	2,304	SEA	LAX	2,623	ATL	LGA	2,325	SFO	EWR	2,697
SFO	LAS	2,194	LGA	ORD	2,528	SEA	SFO	2,216	BOS	SFO	2,591
ATL	LGA	2,099	BOS	SFO	2,513	LGA	ORD	2,175	LGA	ORD	2,475
DTW	FLL	2,008	DFW	LGA	1,902	DEN	SEA	1,934	DEN	LAX	2,145

Models for four roundtrip flights:

ATL → LGA:

$$itinerary\ fare = 422.762 + 37.202 \times distance - 22.074 \times passengers + \varepsilon_1$$

SEA → LAX:

$$itinerary\ fare = 285.505 + 2.590 \times distance - 16.536 \times passengers + \varepsilon_2$$

ORD → LGA:

$$itinerary\ fare = 353.052 + 12.474 \times distance - 35.107 \times passengers + \varepsilon_3$$

SFO → LAX:

$$itinerary\ fare = 268.211 + 0.8799 \times distance - 14.0611 \times passengers + \varepsilon_4$$

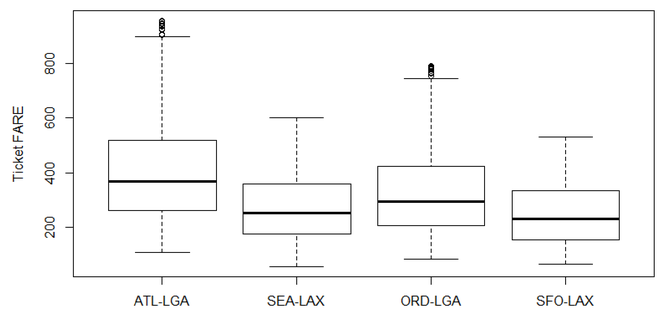
Variables:

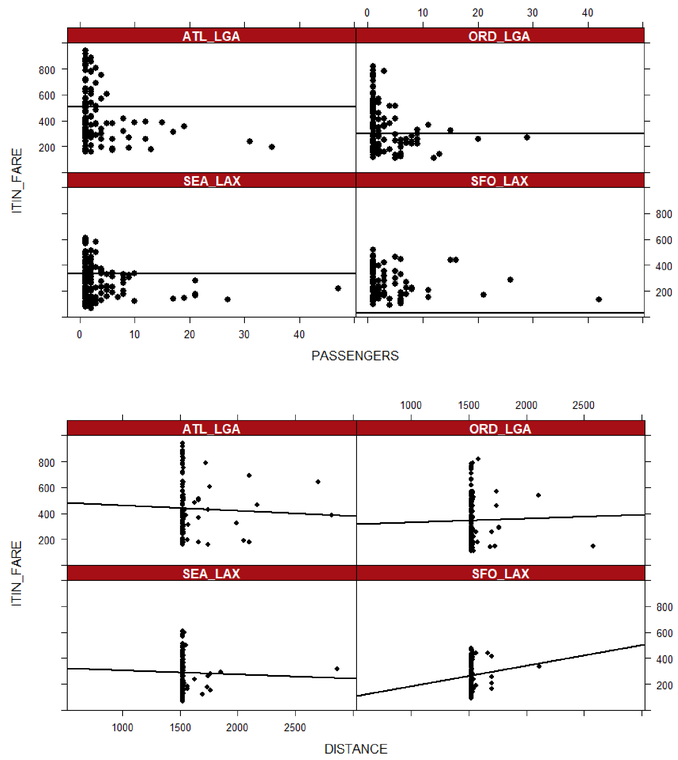
- *Passengers: number of passengers that purchased tickets together*
- *Distance: itinerary distance (including ground transport) in miles*
- *Itinerary fare: roundtrip price per person*

Top 4 Trips in 2017	p-value for F Test
ATL → LGA	2.513e-13
SEA → LAX	7.143e-05
ORD → LGA	3.46e-13
SFO → LAX	0.0002526

All of the p-values are less than 0.1 so the regression models using two variables, Passengers and Distance, are significant at the 0.01 level. Using the regression model with variables Distance and Passengers to estimate the flight fare is better than using the mean of flight fares.

The following box plots show the average fares of the four most popular roundtrip flights in 2017.





Discussion & Conclusions

Contributions:

- Distance and Passengers are both significant values in determining roundtrip flight prices
- Using coupons did not add any significance to the model

Suggestions for Future Research:

- Using data that spans over multiple years could lead to other significant variables
- Another way to study this data would be to look at how far in advance customers bought tickets and at what price to determine if there is a best time to buy tickets prior to a trip
- Adding additional variables such as reward mileage usage or the difference in seat class could lead to different results

Reference

"OST_R | BTS | Transtats." BTS, www.transtats.bts.gov/Tables.asp?DB_ID=125&DB_Name=Airline Origin and Destination Survey (DB1B)&DB_Short_Name=Origin and Destination Survey.

Acknowledgement

University of Dayton Department of Mathematics